Free-Form Variational Inference for Gaussian Process State-Space Models

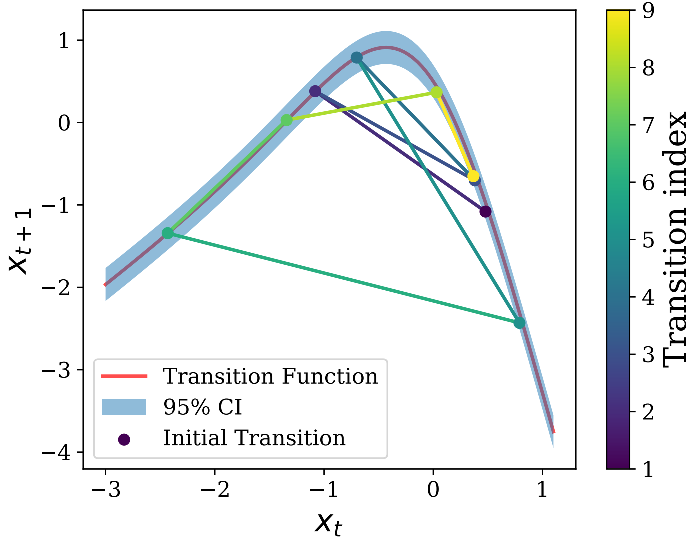
Xuhui Fan, **Edwin V. Bonilla**, Terence J. O'Kane and Scott A. Sisson

International Conference on Machine Learning (ICML)

Honolulu, July, 2023

Free-form Variational Dynamics (FFVD)

- State-space models (SSMs)
  - Modelling temporal data
  - Characterise *dynamics* of a latent state
    - Transition function and observation model



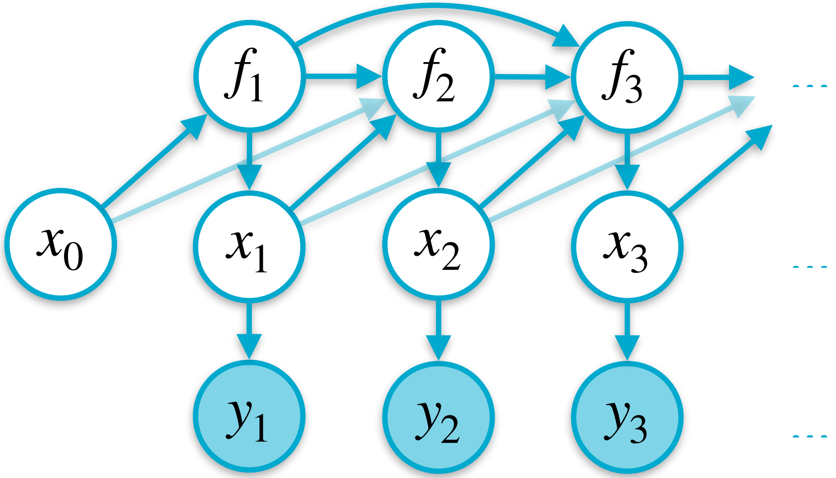
- Gaussian process (GP) SSMs
  - GP prior over transition function
  - Challenges in posterior estimation
    - Non-linear transitions
    - Large number of latent variables
    - Strong dependencies across states
    - Cubic time complexity

FFVD

- ☑ Scalable inference
- ☑ Flexible posterior
- ☑ Collapsed inference
- ☑ State-of-the-art performance

Gaussian Process State-Space Models (GPSSMs)

- Given time series observations $\mathbf{y}_{1:T}$
  - Infer corresponding *latent* states $\mathbf{x}_{1:T}$
- Generative process
  - $\mathbf{x}_0 \sim p(\mathbf{x}_0)$
  - $f(\mathbf{x}) \sim \mathcal{GP}(m_f(\mathbf{x}), \kappa_f(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}))$ (GP hyper-parameters: $\boldsymbol{\theta}$)
  - $\mathbf{f}_t := f(\mathbf{x}_{t-1})$
  - $\mathbf{x}_t \mid \mathbf{f}_t \sim \mathcal{N}(\mathbf{x}_t; \mathbf{f}_t, \mathbf{Q})$ (process noise: $\mathbf{Q}$)
  - $\mathbf{y}_t \mid \mathbf{x}_t \sim p(\mathbf{y}_t \mid \mathbf{x}_t, \boldsymbol{\phi})$ (likelihood parameters: $\boldsymbol{\phi}$)



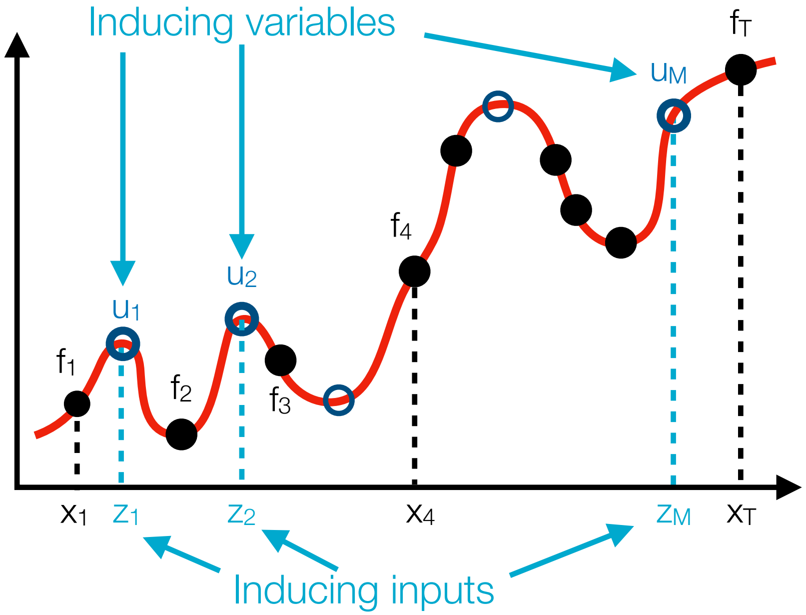
Joint distribution $p(\mathbf{y}_{1:T}, \mathbf{x}_{0:T}, \mathbf{f}_{1:T})$

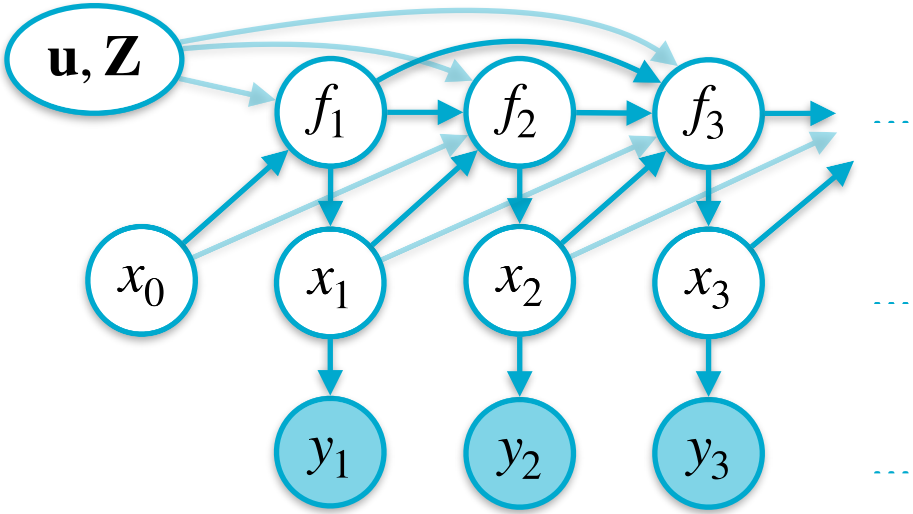
$$p(\mathbf{x}_0)\prod_{t=1}^{T} p(\mathbf{f}_t \mid \mathbf{f}_{1:t-1}, \mathbf{x}_{0:t-1})p(\mathbf{x}_t \mid \mathbf{f}_t)p(\mathbf{y}_t \mid \mathbf{x}_t)$$

Unscalable inference $p(\mathbf{x}_{0:T}, \mathbf{f}_{1:T} \mid \mathbf{y}_{1:T})$

Sparse GPSSMs via Inducing Variables

- Augment the model with *M* inducing variables $\mathbf{u} := \{u_i\}_{i=1}^{M}$ and corresponding inducing inputs $\mathbf{Z} := \{\mathbf{z}_i\}_{i=1}^{M}$





Joint $p(\mathbf{y}_{1:T}, \mathbf{x}_{0:T}, \mathbf{f}_{1:T}, \mathbf{u} \mid \mathbf{Z})$

$$p(\mathbf{u} \mid \mathbf{Z}) p(\mathbf{x}_0) \prod_{t=1}^{T} p(\mathbf{f}_t \mid \mathbf{f}_{1:t-1}, \mathbf{x}_{0:t-1}, \mathbf{u}, \mathbf{Z}) p(\mathbf{x}_t \mid \mathbf{f}_t) p(\mathbf{y}_t \mid \mathbf{x}_t)$$

Estimate posterior $p(\mathbf{x}_{0:T}, \mathbf{f}_{1:T}, \mathbf{u} \mid \mathbf{y}_{1:T}, \mathbf{Z})$

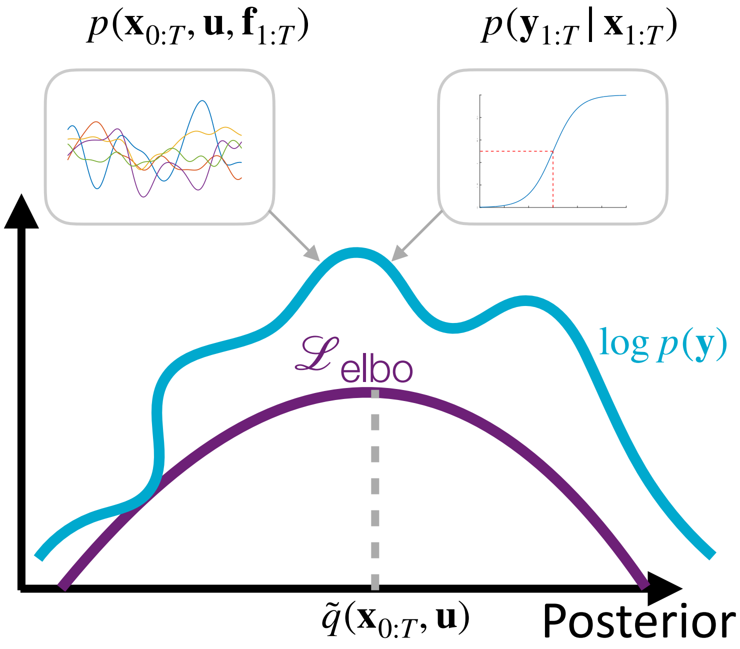
Free-Form Variational Inference

1. Define approximate variational posterior

$$q(\mathbf{x}_{0:T}, \mathbf{u}, \mathbf{f}_{1:T}) := q(\mathbf{x}_{0:T}, \mathbf{u})q(\mathbf{f}_{1:T} \,|\, \mathbf{x}_{0:T}, \mathbf{u}),$$ where

$$q(\mathbf{f}_{1:T} \,|\, \mathbf{x}_{0:T}, \mathbf{u}) := \prod_{t=1}^{T} p(\mathbf{f}_t \,|\, \mathbf{f}_{1:t-1}, \mathbf{x}_{0:t-1}, \mathbf{u}, \mathbf{Z})$$

2. Maximise the evidence lower bound (ELBO) wrt $q(\mathbf{x}_{0:T}, \mathbf{u})$
   - "Optimal" $\log \tilde{q}(\mathbf{u}, \mathbf{x}_{0:T})$ can be obtained analytically
3. Feed optimal log posterior into favourite (SGHMC) sampler
   - Energy function $U(\mathbf{x}_{0:T}, \mathbf{u}) = -\log \tilde{q}(\mathbf{x}_{0:T}, \mathbf{u})$



- Similar computational cost to previous work $O(M^2T)$ but posterior much more flexible

Benefits of Free-Form Variational Dynamics (FFVD)

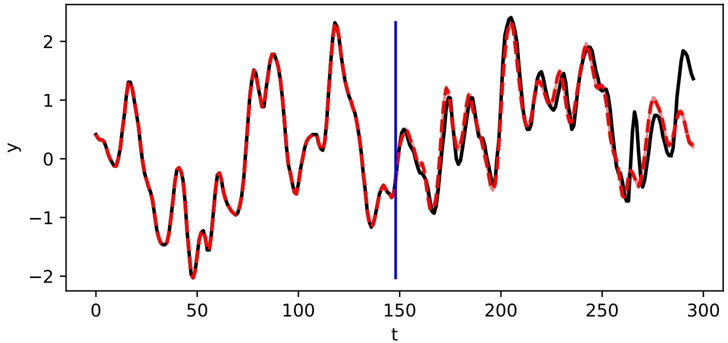
	VGPSSM	IGPSSM	PRSSM	VCDT	FFVD
Coupled $q(\mathbf{x}_{0:T}, \mathbf{u})$	❌	❌	✅	✅	✅
Unconstrained $q(\mathbf{x}_{0:T} \,\vert\, \mathbf{u})$ or $q(\mathbf{x}_{0:T})$	✅	❌	❌	❌	✅
Unconstrained $q(\mathbf{u})$	✅	❌	❌	❌	✅

- Collapsed algorithm
  - Under our factorisation, inducing variables can be integrated out analytically
    1. Run SGHMC using $U(\mathbf{x}_{0:T}) = -\log \tilde{q}(\mathbf{x}_{0:T})$ to obtain $\{\mathbf{x}_{0:T}^{(s)}\}$
    2. Use optimal closed-form expression to sample from conditional $\{\mathbf{u}^{(s)} \,\vert\, \mathbf{x}_{0:T}^{(s)}\}$
  - Better convergence

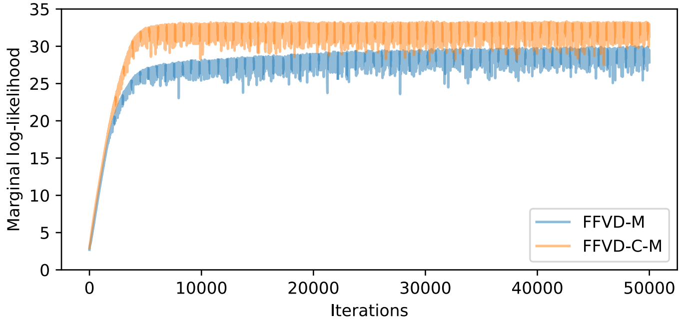
Experiments & Results

- Synthetic data and real systems identification benchmarks with $d_x = 4$

Good test performance



Better convergence of collapsed algorithm



Test RMSE

Methods	Actuator	Ballbeam	Drive	Dryer	Flutter	Furnace
LSTM	0.586 ± 0.411	0.027 ± 0.023	0.537 ± 0.108	0.115 ± 0.029	0.912 ± 0.562	1.261 ± 0.610
VGPSSM	0.580 ± 0.274	0.073 ± 0.011	0.722 ± 0.087	0.241 ± 0.023	1.482 ± 0.218	1.115 ± 0.358
PRSSM	0.497 ± 0.381	0.059 ± 0.013	0.813 ± 0.101	$\mathbf{0.017} \pm 0.042$	1.371 ± 0.156	1.243 ± 0.407
VCDT	$\mathbf{0.239} \pm 0.040$	0.011 ± 0.002	0.585 ± 0.017	0.142 ± 0.003	1.782 ± 0.324	1.166 ± 0.011
FFVD-M	0.358 ± 0.242	0.019 ± 0.018	0.673 ± 0.207	0.205 ± 0.313	$\mathbf{0.280} \pm 0.193$	0.571 ± 0.185
FFVD-C-M	$\underline{0.259} \pm 0.209$	$\mathbf{0.009} \pm 0.011$	0.775 ± 1.615	$\underline{0.065} \pm 0.112$	0.663 ± 0.189	$\mathbf{0.548} \pm 0.051$
FFVD-P	0.388 ± 0.087	0.199 ± 0.045	$\mathbf{0.342} \pm 0.057$	0.317 ± 0.050	0.562 ± 0.088	0.669 ± 0.174

Poster # 708
Session 1
Tue 25th July